КОНСОРЦИУМ ЭКОНОМИЧЕСКИХ ИССЛЕДОВАНИЙ И ОБРАЗОВАНИЯ - РОССИЯ И СНГ
ECONOMICS EDUCATION AND RESEARCH CONSORTIUM – RUSSIA AND CIS

G. Kolodyazhny and A. Medvedev

Microstructure of Russian stock market and profitability of market making

Final report

2002

Abstract

In this paper we study two different issues related to the microstructure of Russian stock market: profitability of market making (major one) and relationship between brokers and their clients. We use data on all the transactions in Moscow Interbank Currency Exchange or a period of more then one year. Our main conclusion is that there are no costless opportunities to get better chances of profiting by engaging in market making. We also found statistically significant relationship between market activity of different customers of the same broker as compared to other participants. We supposed that typical financial firm tends to hide its investments by carrying out market activity via affiliated companies, which naturally become its customers.

Introduction

This paper focuses on the microstructure of Russian stock market. The research was initially motivated by the availability of high frequency data that authors collected while working in the Central Bank of Russian Federation. Our data cover all the transactions that were carried out in the leading Moscow stock exchange (MICEX) over a period of more then one year. It would be certainly impossible within one project to take full use of the unique information we possess, so we focused on the issues that have not been extensively studied in the empirical literature so far.

Russian stock market is rather narrow financial market with one security strongly dominating the others in trade volume. This security, which is the common share of Russian energy monopolist RAO UES, accounts for about 85% of all transactions in MICEX. This makes the data even more attractive since actual market resembles single security market widely assumed in theoretical modeling and measurement issues become less problematic. First issue that we study in the paper and which is the major one is functioning of market makers. In the theoretical literature on market microstructure it is common to assume that market makers compete and do not receive excess profits from their activity. This is questionable if we think of reality. In this paper we analyze the profit opportunities of market makers in MICEX. Our main conclusion is that there are no costless opportunities to get better chances of profiting by engaging in market making in MICEX.

Another issue that we touch upon in the paper is the relationship between brokers and their clients. This is separate issue, which is not related to the first one, but interesting on its own. The fact that, on average, a broker acts on behalf of only about 9 clients during trading session suggests that most brokers are not exactly brokers in a sense that they perform brokerage services for multiple clients. We found statistically significant relationship between market activity of different customers of the same broker as compared to customers of other brokers. This points out to existence of the relationship between broker and their clients. We concluded that typical financial firm tends to hide its investments by carrying out market activity via affiliated companies, which naturally become its customers.

The rest of the paper is organized as follows. In the next section we give a brief focused description of the literature on market microstructure. Then we provide a description of database we have and the way it was constructed. We also provide useful comparison of stock exchanges and shares traded, which seems to be interesting per se. Then we present some preliminary observations and the empirical analysis of the profitability of market making. Finally, we describe the results of the test on the relationship between brokers and their clients.

Market microstructure: a brief review of the literature

Study of market microstructure is an attempt to look inside the "black box" of price formation. Traditional approach in economics is to assume that markets are cleared. The way the equilibrium is achieved is neglected and the main focus is on the equilibrium quantities. Market microstructure literature is concerned with the process of trading, behavior of traders and implications. Nice review of the literature is given in the book O'Hara (1995) and here we give a brief account of the main features with the aim to provide a picture of academic view on the functioning of the financial markets.

The terminology used in the literature is not uniform and varies across papers. Nevertheless, we can provide some commonly used definitions. A *dealer* is a market participant who acts (trades) on his own. A *broker* is a market participant who acts on behalf of its customers. Broker can also be dealer that is he can trade with his own money. A *market maker* or specialist is a market participant who intermediates between buyers and sellers. Market makers provide bid/ask quotes on traded assets and by doing this they maintain the liquidity of the market. One of important purposes of market makers is to solve the problem of time mismatching between order flows.

Market makers can be assigned official status like in the case of Russian T-bills market, where only selected number of dealers were technically able to set long-standing sell and buy orders. Such a market is called specialist market. Each transaction involved a market maker as a counter party. Alternatively, a market can be organized in such a way that every dealer can set orders that are recorded in the order book if not immediately executed. In such a market, trading occurs as a result of matching of sell and buy order flows without involving predetermined market makers. Russian stock exchanges are examples of this type of the market. Although there is no market makers *de juro*, they exist *de facto*. These are those market participants that actively trade in the market having on average no long and short positions in traded securities.

Glosten and Milgrom (1985), Easley and O'Hara (1987) were among first who provided realistic models of market makers' behavior. These models belong to so called information-based models of market microstructure. The main feature of theses models is the existence of informed and uninformed (liquidity) traders. If market maker trades against informed trader then he loses. This loss should be compensated by gains from trades with uninformed traders. Market maker knows that some share of aggregate buy/sell flows is generated by informed investors. The other share is due to uninformed traders who may trade, for example, for liquidity reasons (not related to fundamental information). This allows him to learn some fundamental information from the aggregate order flow he observes. Kyle (1985) further enriched the model by allowing informed traders to fully explore their advantage. The model then takes form of strategic game between market makers and informed investors.

Theoretical models of market making assume that market makers make zero expected profit (under risk neutrality) due to competition. In this paper we will be mainly concerned with evaluating profitability of market making in the leading Moscow stock exchange.

Data

Our database covers all the transactions that were being executed in Moscow Interbank Foreign Exchange (MICEX) between August 4, 2000 and October 24, 2001. It contains slightly less then 6 million rows and has the size of 825 MB. Before we go over to the description of the database we present a brief description of Russian stock market and the role of MICEX.

Russian stock market is divided into organized and unorganized segments, and also ADR market. Organized market consists of a number of stock exchanges, of which two of them (RTS and MICEX) account for more then 95% of total trade turnover. The share of RTS has been steadily declining during recent time for several reasons. Currently MICEX accounts for almost 85% of trade turnover in the organized market. The total value of transactions carried out in MICEX reached RUR 72bn (USD 17bn) in 2000, which is six times as much as in 1999.

The leading position of MICEX is to a large extent due to technological advantages that enable members of the exchange to establish remote trading desks for their clients. By the mid of 2001 about 130 brokers established remote trading desks and the share of trades made via Internet has already reached 40% in money value and 60% in units. It is certain that MICEX is a leader in the organized market with maximum number of transactions (up to 55 000 a day) and that the largest share of domestic and foreign investors trade in MICEX.

The asset structure is also fairly non-uniform. The unambiguous leaders are common shares of RAO UES – the energy monopolist –, which accounts for more then 80% of trade turnover (see table 1). Leading Russian oil company Lukoil holds only 5% of market turnover.

Table 1. Leading stocks: turnover structure

	RAO UES	Lukoil	Surgutneftegaz	Rostelecom	Sberbank
2000					
January	82.12	1.81	1.22	0.39	3.94
February	81.21	8.36	0.66	1.21	2.99
March	76.82	8.15	1.26	1.81	3.07
April	79.59	4.69	1.96	3.5	3.71
May	83.07	6.49	2.52	2.43	1.6
June	89.89	4.26	1.24	1.17	0.99
July	87.18	4.13	1.15	2.69	1.69
August	87.35	3.91	1.07	2.56	1.3
September	84.8	5.49	1.38	2.7	0.59
October	85.36	3.47	2.48	2.24	0.52
November	76.95	8.13	2.31	1.82	0.35
December	84.4	3.08	2.57	1.43	0.84
Total	**83.45**	**5.17**	**1.66**	**2.04**	**1.61**
2001					
January	81.65	7.46	2.19	1.29	0.5
February	89.85	2.79	2.44	0.99	0.4
March	89.99	2.03	1.92	0.91	0.8
April	90.63	2.13	1.55	1.65	0.63
May	81.36	5.4	3.45	2.38	0.62
June	71.19	8.99	4.49	3.29	0.93
Total	**84.62**	**4.58**	**2.62**	**1.71**	**0.64**

The authors had collected the data when working as economist in the Central Bank of Russian Federation. Initially, data were stored in text files. For each trading day in the sample we had a file containing information on all the transactions and a file containing all the orders submitted during trading session. The total number of files was about 500 and the size of the collection of files was about 2 GB. Each transaction was represented by a row in a file with the following information:

- Transaction number;
- Time (with the precision of one second);
- Name of a stock (For example, RAO EES)
- Trade regime (stocks of tier 1, stocks of tier 2, out-of-listing stocks, regime of negotiated transactions);
- ID of the stock (for example, for RAO EES – RU0008926621);
- Direction of the transaction (B – buy, S – sell);
- Broker's account;
- Name of the broker;
- Price of the transaction;
- Number of lots of securities (For RAO EES 1 lot = 100 stocks);
- Volume of the transaction;
- ID of the broker of the counter party;
- Name of the broker of the counter party;
- A Remark

Text format is not suitable for calculations so the program has been written to convert text files into MS Access format. In fact, that was the most difficult part of the research. First, it required developing software to process text files, second, processing was quite time consuming. For example, conversion of one-month data required a couple of days.

The major problem that is encountered by many researchers is that stock data do not allow distinguishing between different investors. Financial institutions typically operate both as investment funds that attract and manage capital and as brokers that provided intermediation services. Typically stock exchange data allow separating brokers from their clients but not one client from another one. Fortunately, our data provide means to overcome this problem. This became possible due to the presence of so called remarks in the transaction statistics. Brokers tend to use remarks in the "order window" to determine which client sent the request. These remarks cannot be classified since they are made on discretion, which, however, is not necessary to do in order to distinguish clients.

As it was pointed out, common stocks of RAO UES account for about 85% of all the transactions. This means that if we restrict our analysis to transactions with this leading stock, we will not loose much information about market microstructure. However the gain will be substantial, since it always easy to deal with one stock and one price then with a range of stocks and prices. All the analysis that will be presented in following sections is computed on the basis of data that correspond only to common stocks of RAO UES.

Preliminary observations

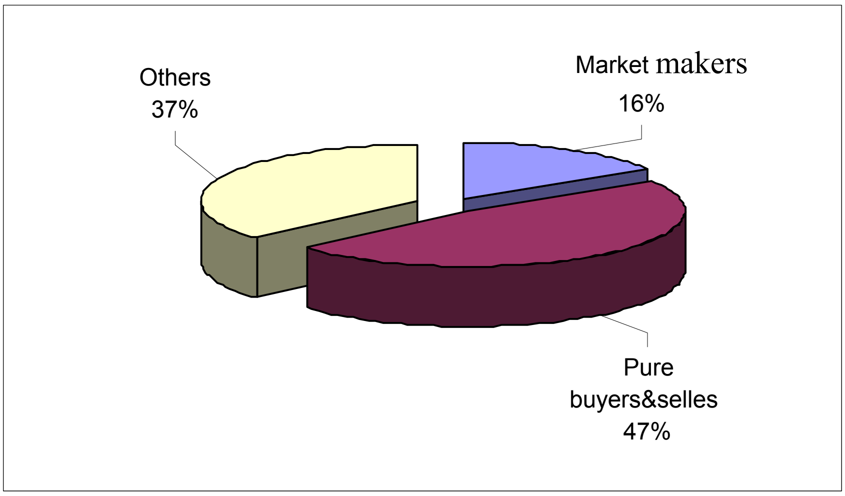
In this section we present some preliminary findings that will provide a general picture about the microstructure of the stock market under consideration. As it was already mentioned in the previous section, we will focus on single company RAO UES that accounts for more then 85% of all transactions in MICEX both with respect to the number and the volume.

On practice, it is difficult to classify traders into market makers, informed and uninformed. What we observe in reality is number of shares bought and sold by every trader during a day. By definition, a market maker stays neutral on average, that is he sells and buys equal number of shares on average. This feature provides us with the means to pin down market makers using data on transactions. This approach seems to be problematic in our case for number of reasons. The main difficulty is that there are no pure market makers that profit solely from intermediation between buyers and sellers. Second, it is rather difficult to access their profits if they are not perfectly neutral on the daily basis. In this paper we will take a simplified approach by considering only traders that remained perfectly neutral (sold and bought the same number of shares) during a day as marker makers. This definition, as well as the other ones introduced below, is day-dependent meaning that an investor can be classified as market maker in one day and be classified as, for example, buyer in the other day. An investor is called buyer (seller) if the number of shares he bought exceeded the number of shares he sold. An investor is called pure buyer (pure seller) if he was only buying (selling) shares in a given day.

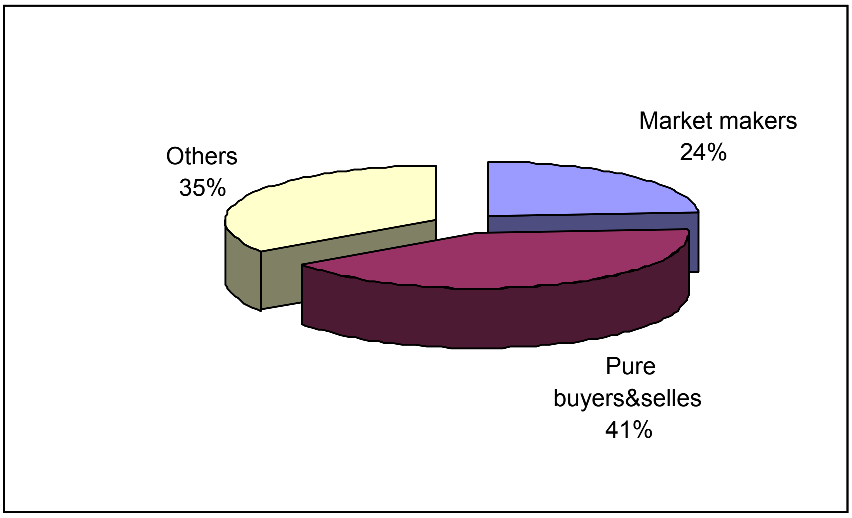
The immediate criticism of our classification is that we could possibly enlarge the pool of market makers by including those who were "almost" neutral during a day. This option has been analyzed and we find no satisfactory way to define the notion of "almost". More importantly, the definition of market maker is subject to potential selection bias. Indeed, an investor might not be willing to sell all the shares he bought during a day if it leads to realized losses. To defend our classification, we note that market makers we cover in our classification are probably those that extensively use leveraged borrowing to increase the scale of potential profits. For them, the only way to pay out is to close their position in the end of the day whatever net effect is. Below we provide some empirical facts that suggest absence of selection bias.

Pic. 1 and 2 summarize the average daily structure of market participants according to our classification by number of investors and volumes of transactions (turnover):

Pic. 1 Structure of market participants (number of daily transactions)



Pic. 2 Structure of market participants (volume of daily trade)

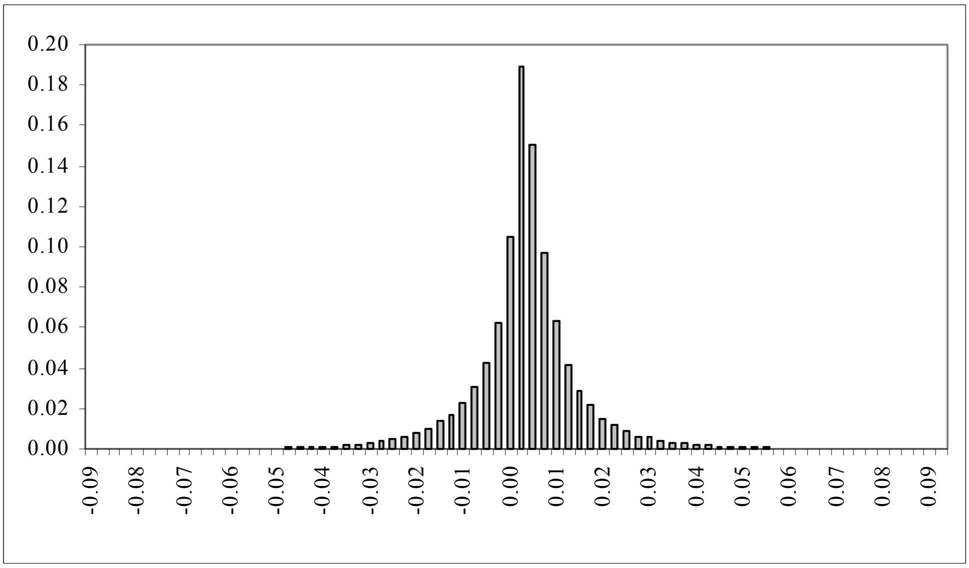


As it follows from these diagrams, on average, 16% of market participants are comprised of market makers (as we defined them) with about 50% being pure buyers or sellers. In volume terms, market makers account for almost quarter of the total turnover. This figure is not so impressive as one would expect from the role of market makers in the market. This is an obvious draw back of our classification.

Nevertheless, even given quite restrictive definition of market makers, we are left with significant share of the trade they account for.

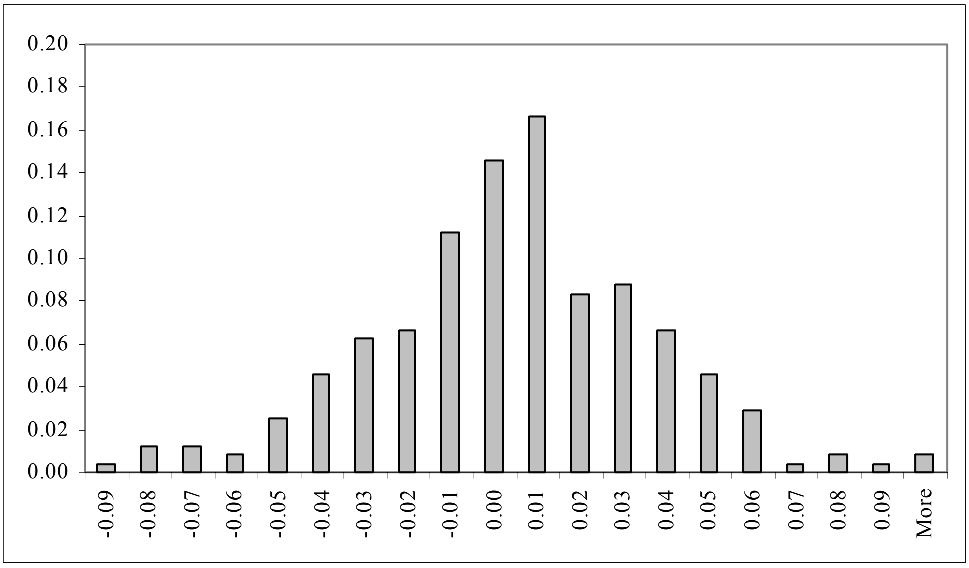
From our data, we are able to estimate the *ex-post* performance of market makers. The sample distribution of the daily rate of return of market making is shown in pic. 3. As a benchmark, we draw the distribution of daily return from the simplest (myopic) strategy of buying in the beginning and selling in the end of a trading day[1] (pic. 4).

Pic. 3 Daily rate of return from market making



Sample mean: 0.18%; standard deviation: 0.99 percentage points.

Pic. 4 Daily return from the myopic strategy



Sample mean: -0.02%, standard deviation: 3.28 percentage points

[1] We could consider just opposite strategy but it would lead us to the same conclusions.

The return from the myopic strategy is, in fact, the market return. Therefore, judging from the estimates of two first moments for both distributions, we conclude that market making results in excess return (even of higher order in percentage terms) with lower associated risk. This observation suggests that there are profitable opportunities for market making in the Russian stock market.

Let us evaluate the profitability of market making with single parameter, which is the average probability of success of a representative market maker over the period of observations. First we calculate the share of market makers that made profit in a given day, and then we take an average over the whole period. It is straightforward to classify market making between profitable and non-profitable by the difference between the value of shares sold and the value of shares bought. Recall, that according to our classification market makers buy and sell equal number of shares in a given day.

This procedure gives an estimate of the arithmetic average of daily probabilities. Due to sufficient degree of aggregation it has approximately normal distribution with variance that we can estimate the following way. Denote as π_t the true (underlying) probability of success in day t. Let:

$$\pi = \frac{1}{T}\sum_t \pi_t$$

Then if we observe N_t speculators then the variance of estimate of π_t is equal to:

$$V(\tilde{\pi}_t) = \frac{\pi_t(1-\pi_t)}{N_t}$$

and the variance of the estimate of the average probability is:

$$V(\tilde{\pi}) = \frac{1}{T^2}\sum_t \frac{\pi_t(1-\pi_t)}{N_t}$$

This variance can be evaluated by substitution of true probabilities by their estimates. As a result we obtained π=0.657±0.002.

For comparison, the estimated probability of success from myopic strategy is only 0.508.

It should be noted that it is not perfectly correct to use this characteristic as a reliable indicator of superiority of one distribution over another. Indeed, it may be the case that probability of success is greater but the expected utility, which determines correct ordering, is lower. In fact, it is also not strict to work only with first two moments of the distribution. Nevertheless, we believe that probability of incorrect inference is negligible if compared to the extent to which the use of single parameter simplifies analysis.

Before we proceed with the discussion of these finding, it is important to assure that there is no sample bias so that results are not spurious. By our classification, market makers are those investors that were perfectly neutral by the end of the day. The main criticism is that the sample is biased towards more successful market makers because investors may be unwilling to realized losses by closing up their positions. Let us try to test indirectly if indeed there may exist selection bias in our estimates. If the story about investors non-wiliness to fix losses is correct then we expect to find that in days with high average probability of success, the number of market makers (according to our classification) is larger then in days with low probability of success. Also we should double-check this in the following way. Since short selling is not always available then investors that do not want to realize losses are likely to remain among buyers in a given day. Hence there should be negative correlation between the number of market makers and the number of buyers.

Since we are interested in a short-run correlation then it is reasonable to take first differences in data. Instead of the number of buyers we considered the difference between the number of buyers and the number of sellers. This was done to exclude possible common factors that affect market activity. The results are presented in table 2.

Table 2 Correlation matrix

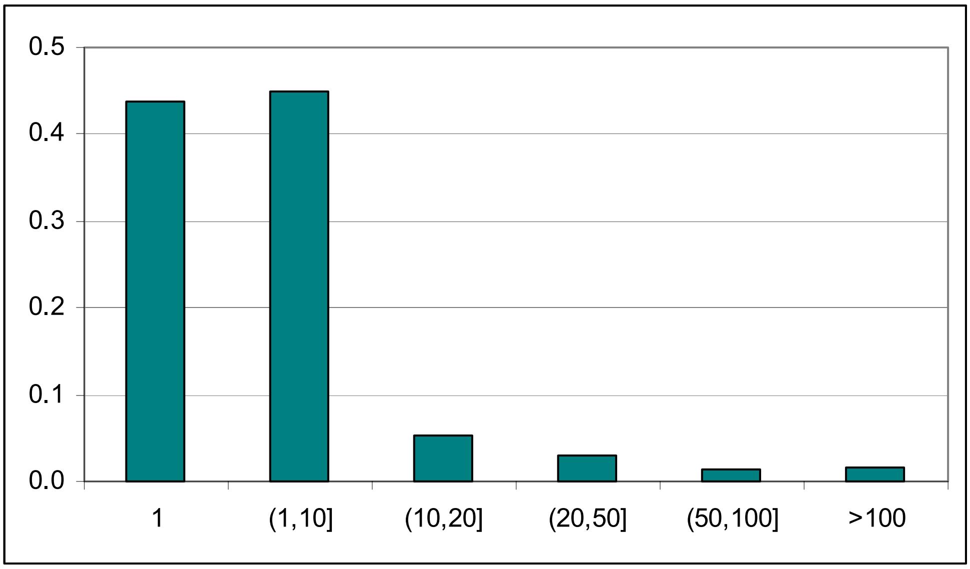
	#market makers	%success	#buyers-#sellers
#market makers	1.00	0.27	-0.03
%success	0.27	1.00	-0.20
#buyers-#sellers	-0.03	-0.20	1.00

According to the argument, in “bad” times more investors face losses and might not be willing to close up their positions. If “bad” times are correctly measured by the average probability of success then we expect to find positive correlation between the success probability and the number of market makers on one hand, and negative correlation between the number of buyers and the success probability on the other hand (due to short-selling difficulties). Although the results are in accord with the expectations, the correlation of 0.27 is not very convincing. Moreover we are surprised to see no significant correlation between the number of buyers and the profitability of market making. These observations allow us to conclude that if any, the sample bias should not be significant.

Our main findings so far suggest that there are highly profitable opportunities for market making in the Russian stock market. Does this mean that this market segment is not competitive as is always assumed in the theoretical model of market making? This question is not easy to ask because we implicitly made quite unrealistic assumption that market makers are homogeneous. Suppose that new market maker enters the market. Can we claim that he will immediately face highly profitable opportunities? Our story of market making obviously lacks some learning aspects that are necessary to make it realistic. Learning market is costly so there may be no “free lunch”. The distribution of the rate of return from market making (pic. 1) reflects opportunities faced by an average (with respect to knowledge) market maker and there is no reason to believe that newcomer will have the same odds. One way to enrich our analysis, which we will pursue in the next section, is to allow for two types of market makers: so called professionals and amateurs. Then our task will be to assess opportunities faced by representative amateur, which will give more realistic picture of how attractive is entering the business of market making. To simplify the analysis, we will assess the attractiveness by single parameter – probability of success, with benchmark of 50% (approximately the same as in the case of myopic strategy).

We finish this section by presenting another interesting piece of information about market microstructure. MICEX is organized in such a way that only limited number of investors can directly participate in the trade. These investors are called brokers. In this connection, it might be interesting to observe distribution of the number of clients per broker. The average number of brokers that traded daily is 286, whereas the average number of participating investors is equal to 2554. This observation suggests that a broker has, on average, about 9 clients (including himself). In pic. 5 we present the average daily distribution of the number of clients per broker computed for the whole period of observations.

Pic. 5 Average distribution of the number of (active) clients per broker



This diagram was constructed the following way. We chose six intervals for the number of clients per broker. Then for each day and for each interval we counted brokers that acted on behalf of clients, number of which falls within corresponding interval. Hence for each day we have six numbers. Then we summed these numbers across all the days and scaled them to have final values summing up to one.

The picture suggests that a significant share of brokers make operations on behalf of only one client (most probably themselves). Most of the others have no more then 10 clients. There are several brokers that serve numerous clients (more then 100). It seems reasonable to expect that brokers that have few clients have close relationship with them. On the other hand, large-scaled brokers (with respect to the number of clients), most probably, are not directly related to their clients. However, one might expect that they provide expert advise to their clients, which we call as indirect relationship. The issue of the relationship between clients and brokers will be analyzed later in the paper.

Profitability of market making

Preliminary findings discussed in the previous section suggest that there are attractive opportunities for market making in the Russian stock market. We found that market making generates higher order returns with lower associated risk, as measured by standard deviation, compared to the benchmark. As it has been noted, we implicitly assumed homogeneous market makers. Obviously, market making involves costs of learning the market, which should not be neglected when assessing the attractiveness of market making opportunities. These costs cannot be measured directly. An alternative way is to drop assumption of homogeneous market makers and postulate existence of two types of market makers: experienced (professionals) and inexperienced (amateurs). Then the profit opportunities faced by inexperienced market makers is the right measure of attractiveness of this activity. In this section we develop a statistical model that will then be used to estimate profit opportunities of two types of market makers. To simplify matters we will measure profit opportunities by the probability of success. Since the share of experience market makers is not known, we will obtain only 95% confidence area in the two dimensional space with the probability of success of professionals and amateurs as coordinates.

In order to get an idea of how different are investors with respect to their success probabilities we will make use of time series property of our data. If investors are homogeneous then the conditional probability of success of an investor is independent of his results in the previous day. Indeed, otherwise his current success is an informative signal for his future success, which can be the case only if investors have different unconditional (prior) probabilities. Hence the reasonable approach is to consider adjacent trading days and estimate conditional probabilities of success.

This approach has a number of difficulties, which become evident once we try to implement it. First, on average, only one third of speculators that participated during some trading day also participate during the next day. Apart from reduced coverage, we may potentially have problems with sampling bias. Indeed, we found that the average probability of success computed over reduced sample is equal to 0.643 ± 0.004 instead of 0.657 ± 0.002 that we had before. Given the estimated standard error, we conclude that the sampling is not perfect. This curious result is difficult to explain, which, in fact, is not so important given that the new estimate is not (economically) significantly different. So we will simply stick to it instead of the old one.

Second, the conditional probability of success of an investor for some day given that he participated also during the next day might be higher then that for an "average" investor. This bias may emerge if success in one day increases probability of participation in the next day. Indeed, the estimated mean of the probability of success of an investor conditional on his participation in the previous day is 0.64 whereas conditioning on the next day participation gives 0.68. This difference is indeed significant not only from the statistical point of view but also economically. As it will be seen later, under reasonable assumptions this discrepancy does not change formulas that we obtain for estimation purposes.

Formal setup

Let us formalize the heterogeneity of investors in the following formal setup. We assume that there are two types of speculators in the market: ”professionals” and “amateurs”. First type of investor will be referred as high type and the second as low type. The share of high type is fixed and denoted as λ. Let π_t be the probability of success of an average investor in day t, $p_t^{h(l)}$ – probability of success of high (low) type. Then:

$$\pi_t = \lambda p_t^h + (1-\lambda) p_t^l$$

Let us denote as $p_t^{s|s}$ the conditional probability of success of a representative investor given that he succeeded in the previous day. Suppose that success in one day has equal multiplier on the probability of continuation independent of the type of investor. Then the share of the high type among all speculators that participated today (t) and yesterday ($t-1$) is always λ. It follows that probability of high type given success yesterday is equal to:

$$\Pr(h \mid s) = \frac{\Pr(s \mid h)\Pr(h)}{\Pr(s \mid h)\Pr(h) + \Pr(s \mid l)\Pr(l)} = \frac{p_{t-1}^h \lambda}{p_{t-1}^h \lambda + p_{t-1}^l (1-\lambda)}$$

Similarly

$$\Pr(l \mid s) = \frac{\Pr(s \mid l)\Pr(l)}{\Pr(s \mid h)\Pr(h) + \Pr(s \mid l)\Pr(l)} = \frac{p_{t-1}^l (1-\lambda)}{p_{t-1}^h \lambda + p_{t-1}^l (1-\lambda)}$$

And the conditional probability $p_t^{s|s}$ is equal to:

$$p_t^{s|s} = p_t^h \Pr(h \mid s) + p_t^l \Pr(l \mid s) = \frac{p_t^h p_{t-1}^h \lambda + p_{t-1}^l p_t^l (1-\lambda)}{\pi_{t-1}} \tag{1}$$

Let us assume that probability of success of high type is relatively smooth over time ($p_t^h \approx p_{t-1}^h$). In other words, the volatility of the probability of success of the average investor is accounted for mainly by the volatility of corresponding probability for low type (non-professionals). This assumption seems to be reasonable and also very useful since allows us to solve for p_t^h. Expression (1) can be now written in the following form:

$$p_t^{s|s} \pi_{t-1} = (p_t^h)^2 \lambda + p_{t-1}^l p_t^l (1-\lambda) \tag{2}$$

Also we have:

$$\pi_{t-1} = \lambda p_{t-1}^{h} + (1-\lambda) p_{t-1}^{l} \tag{3}$$

$$\pi_{t} = \lambda p_{t}^{h} + (1-\lambda) p_{t}^{l} \tag{4}$$

The system of equations (2), (3) and (4) can be easily solved to obtain:

$$p_t^h = \frac{\pi_t + \pi_{t-1}}{2} + \sqrt{\frac{1-\lambda}{\lambda}\pi_{t-1}(p_t^{s|s} - \pi_t) + \left(\frac{\pi_t - \pi_{t-1}}{2}\right)^2} \tag{5}$$

This expression relates high type probability to conditional and unconditional probabilities of success, which can be estimated from the data. The issue of estimation is complicated by non-linearity of the relationship. On the other hand we seek estimates of arithmetic averages of fundamental probabilities but not their spot values:

$$p^h \equiv \frac{1}{T}\sum_t p_t^h$$

$$p^l \equiv \frac{1}{T}\sum_t p_t^l$$

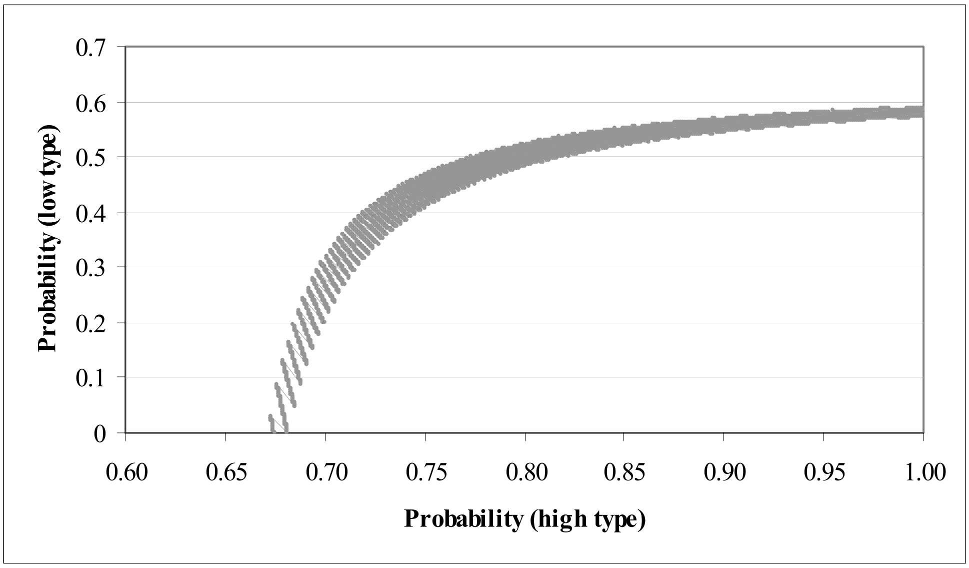
$$\pi \equiv \frac{1}{T}\sum_t \pi_t$$

Given sufficiently long series of data we are able to make assumption about normality of some estimates and this will help considerably in obtaining confidence intervals. The estimation technique is described in Appendix.

Estimation results

The estimation procedure described in the previous section was implemented on our database. First we found estimate of a, then evaluated variance of the estimate and finally determined 95% confidence area using its definition given by (15), (16) and (17). In the last step, we took 100 uniformly distributed values from intervals [-2,2] and (0,1) where t and λ are defined and then plotted those 10000 points (x,y) that satisfy conditions $x \leq 1$ and $y \geq 0$. The resulting plot is shown below.

Pic. 6. The 95% confidence area for true success probabilities



The minimum λ for which the confidence area for underlying probabilities intersects with square $[0,1]^2$ is about 0.13 so we can reject hypothesis that speculators are homogeneous with respect to their probabilities of success. Now we are able to make judgments about possible values of low type probability that is key variable we wanted to estimate. There are two observations in order. First, p^l is always smaller then 0.6, which does not look such optimistic as 0.66. This is the maximum possible value for the low type probability, which can be true only if "professionals" make sure profits. For example, if some investors manipulate the market, then it might be reasonable to expect $p^h \approx 1$. However, this would mean that more then 10% of market makers actually manipulate the market. This looks quite unrealistic. If there is any manipulation, it probably happens occasionally and should not affect aggregate results.

Second observation is that if we believe that low type investors can succeed with more then 50% probability (that is benchmark) then we will have to accept that high type has probability of at least 77%. This figure seems to be exceptionally high, although, there is no strict way to justify our belief. Success probability of 77% implies huge expected return from market making provided that the return distribution is not highly asymmetric[2]. The theory tells us that due to presence of informed traders, market makers infer the fundamental information from trades and earn on uninformed traders. Hence, the market making profit opportunities depend crucially on the unequal distribution of information in the market. In this

[2] Suppose that there is large probability of very small profits and small probability of huge losses. Then the probability of success can be high by the profit opportunities are not especially favorable. This is general observation that use of success probability to measure attractiveness of activity is correct only if return distribution is near symmetric. This is what we observe for the whole sample of market makers (see pic. 1).

paper we focus on the market for a single stock RAO UES. The company represents a huge energy monopolist with uncertain future and highly non-transparent accounting. According to market participants, the market has no clear understanding of bounds for the fundamental value of RAO UES. Under these circumstances, the dynamics of the price of common shares of RAO UES is mainly determined by general political and economic factors rather then company specific ones so that one cannot expect the information to be unequally distributed in the market. As a consequence, market making should not yield huge returns in this segment of stock market.

As is evidenced by pic. 6, professionals have the probability of success of at least 66%, which is still high. This observation presents a puzzle that we are not ready to fully resolve. This might evidence that there is probably some sample bias in the classification we adopted (see discussion above) that increases all the probability estimates. This bias should not be large as is evidenced by the indirect test so that we are inclined to believe that 77% would still be large even if corrected for the sample bias. Consequently, the argument against low type probability being greater then 50% remains valid. To sum up, we conclude that probability of success of amateurs is *unlikely* to be better then 50/50 chances.

Relationship between brokers and their clients

In this section we will test for the existence of the relationship between brokers and their clients. As it is evident from preliminary analysis, most of brokers have only several clients, which are probably directly related to them (affiliated companies). This kind of relationship we call direct relationship. The indirect relationship arises when brokers provide market advise to their clients, for example, in the form of daily market review. Both types of relationship should result in a relatively similar behavior of clients of the same broker. Hence the right way to test the hypothesis of existence of the relationship is to estimate the degree of coordination in actions of investors that use services of one broker as compared to others in the market.

Methodology

The statistical methodology is adopted from the empirical literature on herding behavior (Lakonishok, Shleifer and Vishny (1992), Wermers (1999) among others) but is applied to totally different context. The herding statistics, introduced by Lakonishok, Shleifer and Vishny (1992), measures the degree of concentration of the trade of a group of investors in different stocks. Here we want to measure something very similar. Namely, we would like to measure the degree of coordination between actions of investors that are associated with the same broker. In case of the herding measure we have a group of investors being fixed and stocks being varied. Here we fix the stock but vary groups of investors by considering different brokers.

Let index t refer to some day in the sample and let i refer to a broker, who performed operations on behalf of its clients (or himself) during the day. If an investor increased holdings of the stock during the day then we say that he was a buyer during this day. Similarly we define a seller. Let p_{it} be the share of buyers among all participating clients of broker i during day t. Under null hypothesis p_{it} should be an unbiased measure of general propensity to buy p_t among all investors during day t. There is no a priory reason to believe that it is equal to one half since investors do not trade equal amounts. This propensity is proxied by the actual ratio of buyers among all participating investors during day t. Under null hypothesis the following value is a realization of random variable with zero mean:

$$h_{it} = |p_{it} - p_t| - E|p_{it} - p_t| \tag{18}$$

The expectation on the right-hand side of (18) is computed given that p_{it} has binomial distribution with mean p_t:

$$E|p_{it} - p_t| = \sum_{j=1}^{N_{it}} C_{N_{it}}^{j} (p_t)^j (1-p_t)^{N_{it}-j} \left| \frac{j}{N_{it}} - p_t \right|$$

Here N_{it} is the number of participating clients of the broker i in day t. Let us define aggregate daily measure:

$$H_t = \frac{1}{I_t} \sum_{i=1}^{I_t} h_{it} \tag{19}$$

If the number of brokers I_t is sufficiently large then under null hypothesis H_t is normally distributed with zero mean. Given a series of observations of H_t over the sample period, we may use t-test to test the null hypothesis.

The statistical inference can be also made without assuming that H_t are distributed normally. Instead, we assume that only H has normal distribution, where

$$H = \frac{1}{T}\sum_t H_t$$

Its variance can be estimated from the variances of h_{it}. Under null hypothesis:

$$V(h_{it}) = E\left(|\, p_{it} - p_t\,|^2\right) - E^2(|\, p_{it} - p_t\,|) = \frac{p_t(1-p_t)}{N_{it}} - E^2(|\, p_{it} - p_t\,|)$$

Given assumption on the independence of estimate in tie and across brokers, which has been implicit so far, the variance of H is equal to:

$$V(H) = \frac{1}{T^2}\sum_t \frac{1}{I_t^2} V(h_{it})$$

Now we have an observation of a normal random variable with zero mean and known variance under null hypothesis. The test for the hypothesis is based on evaluating probability of this observation.

The validity of null hypothesis can be also checked visually by means of a simulation. In this paper we performed several simulations of the distribution of H_t using the following methodology. Taking p_t as given we simulated p_{it} assuming binomial distribution and computed H_t. For each day t we simulated 100 values of H_t. The total number of days in the sample is equal to 240, so each simulation consists of 24000 observations of H_t.

Estimation results

The value of H was calculated in several different ways by imposing restrictions on the minimum number of active clients of a broker NC_{min}. These restrictions were imposed to exclude noisy observations of brokers that traded on behalf of very small number of clients. For each day t in the sample, H_t was computed by averaging across only those brokers that performed operations on behalf of at least NC_{min} of their clients (probably including themselves) during this particular day. In table 3 we present estimates of H and t-statistics for a range of values of NC_{min}. The test based on known normal distribution for H also indicated high significance.

Table 3 Results of the test for the presence of relationship between brokers and their clients

NC_{min}	5	10	20	30
H	0.031	0.029	0.021	0.016
t-statistics	29.23	23.67	17.93	14.28
Av. num. of brokers	60	29	15	12

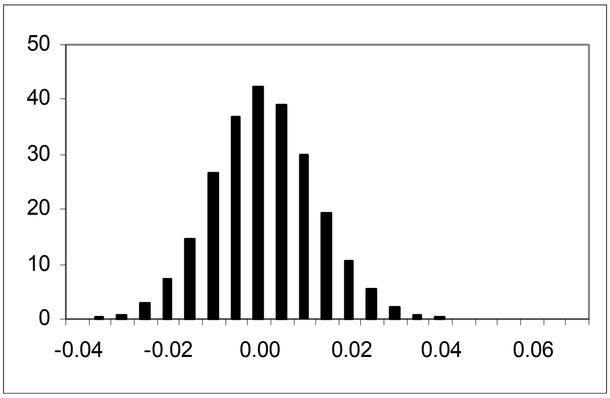
The t-statistics are highly significant, which is usual when a large number of observations is available. The highest value of H is obtained for the case of mild restrictions as indicated by the second column of the table. The value of H can be loosely interpreted as an average bias of the ratio of buyers within one broker as compared to the total value. Here the bias is equal to 3 percentage points, which does not seem to be particularly significant from the economic point view. One possible explanation of this fact is that we did not take into account full information about investors' behavior such as volumes of their transactions. Instead, we used the simplest classification into buyers and sellers. Another possible reason is that we tested the hypothesis on high frequency data. Nevertheless we are able to reject the null hypothesis statistically.

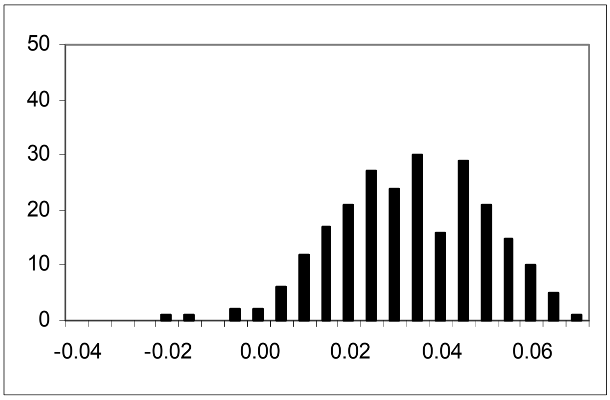
As is evidenced by the last row in the table, the average number of brokers quickly decreases as restrictions on minimum number of participating clients become stronger. For example, on average only 29 brokers performed operations on behalf of at least 10 clients during a day. This fact may also impair the validity of the statistical inference based of t-statistic. Indeed, as it was mentioned in previous section, the test relies on the normality assumption due to sufficient aggregation. We double-checked the results by applying the test based on known normal distribution of H. The estimated standard deviation of H is around 0.0008 for all four cases. The lowest value of H we obtained is 0.016, which is sufficiently different from zero given standard deviation 0.0008.

Another way to check the validity of results is to see visually the difference between actual and null-hypothesis distribution of H_t, we simulated the latter for NC_{min} = 5 and 30.

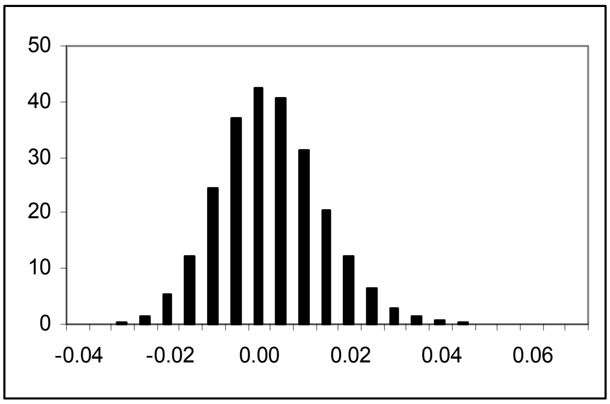
Simulation **Actual data**

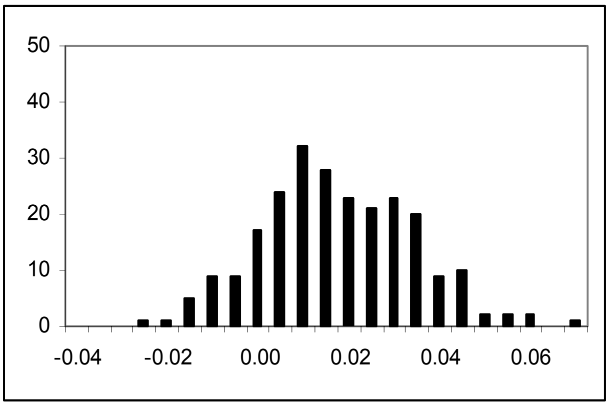
A: $NC_{min} = 5$





B: $NC_{min} = 30$





Our results suggest that there exist a relationship between brokers and their clients. Even for quite large number of clients per broker (here more then 30) we obtained significant estimate of *H*. This means that both direct and indirect relationship exists. As it can be expected the strength of the relationship measured by *H* declines as a restriction on the minimum number of clients becomes stronger, so that the direct relationship is obviously stronger.

Could our results be biased? Yes, if the choice of an investor between large-scaled and small-scaled brokers may be somehow related to the type of strategy he pursues. In this case the difference in behavior between clients of different brokers may be observed if we aggregate *H* across brokers of different size[3]. In particular, this reasoning may explain why we observe decrease in the bias as restrictions on the

[3] Here by size and scale we mean average number of active clients. This may not be related to the size of the broker defined as a value of the clients' portfolio.

number of participating clients become stronger[4]. To make sure that our result are not consistent with this explanation, we performed the test by imposing the restriction that the number of participating clients to be no less then 5 and no more then 10. Under this restriction, we aggregate over relatively homogeneous brokers in terms of their size. The resulting value of H is equal to 0.030 (t-statistic 18.31). Note that these calculations are based on approximately half of the sample used to calculate H value in the second column of table 1. Nevertheless, the new value is only slightly smaller.

The results we obtained in this section represent an interesting piece of information, which, however, is not directly related to what we did in previous sections. We established that brokers and their customers are somehow related. Our data suggest that brokers are not very active dealers and volumes of their trades are surprisingly small. All these observations suggest that there is a tendency to hide investment activity by firms via affiliated companies, which are customers of these firms. This explains the observed "direct" relationship between brokers and their clients. What seems puzzling is the existence of what we call "indirect" relationship. We found that large-scaled brokers also tend to be somehow related to their clients. One possible explanation is that brokers provide market advice to their clients. Alternatively, style of investment may affect the choice of a broker.

Summary and conclusions

This paper presents several empirical results on the market microstructure that were obtained using unique database covering all the transactions in the leading Moscow stock exchange (MICEX) over a period of more then one year. Two issues have been investigated:

- Profit opportunities for market making in the stock market
- Relationship between brokers and their clients

Preliminary calculations showed that the market making allows earning higher order returns as compared to the benchmark with lower associated risk. The probability of success of engaging in market making was estimated to be about 66%, which was deemed to be high. However, careful statistical analysis showed that investors are not homogeneous with some of them having higher probability of success then the others. Assuming existence of two types of investors – "professionals" and "amateurs" – we argued

[4] Notice that we estimate p not from the whole sample, but from the restricted one for the purpose of consistency.

that the probability of success of the latter is at least not higher then 50%. This result suggests that without costly learning there are no excessive gains from market making.

We found statistically significant relationship between market activity of different customers of the same broker as compared to customers of other brokers. This points out to existence of the relationship between broker and their clients. We classified this relationship as direct (case of small-sized brokers) and indirect one (case of large-sized brokers) and found evidence of both types of relationship. We supposed that typical financial firm tends to hide its investments by carrying out market activity via affiliated companies, which naturally become its customers.

References:

Easle, D., and O'Hara, 1987, "Price, trade size, and information in securities markets," *Journal of Financial Economics* 19, 69-90.

Glosten, L., and P. Milgrom, 1985, "Bid, ask, and transaction prices in a specialist market with heterogeneously informed traders," *Journal of financial economics* 13, 71-100.

O'Hara, M., 1995, "Market microstructure theory," Blackwell Publishers Inc., Cambridge MA, USA.

Kyle, A.S., 1985, "Continuous auctions and insider trading," *Econometrica* 53, 1315-1336.

Lakonishok, J, A. Shleifer and R.W. Vishny, 1992, "The impact of institutional trading on stock prices," *Journal of Financial Economics*, 32, 23-34.

Wermers, R., 1999, "Mutual fund herding and the impact on stock prices," Journal of Finance, No. 2, April, 581-622.

Appendix: Estimation technique

Estimation technique

Although we are not able to estimate p^h and p^l we may try to find a subset of space (p^h,p^l), where the true probabilities lie with at least 95% probability given our observations. This set is defined as a union of 95% confidence subsets corresponding to all possible values of parameter λ intersected with the scare $[0,1]^2$.

Let us denote $\delta_t = \left(p_t^h - \frac{\pi_t + \pi_{t-1}}{2} \right)^2$, which is square of the difference between probability of success of high type and average probability of success of representative investors today and yesterday. From (5) it follows that:

$$\delta_t = \frac{1-\lambda}{\lambda}\pi_{t-1}(p_t^{s|s} - \pi_t) + \left(\frac{\pi_t - \pi_{t-1}}{2} \right)^2 \tag{6}$$

Define:

$$\delta \equiv \frac{1}{T}\sum_t \delta_t$$

Let us use the following approximation:

$$p^h \approx \pi + \sqrt{\delta} \tag{7}$$

Note, that all the relationships so far involve only true values for underlying parameters. Now we go over to estimation issues.

Let us first discuss how to estimate δ for arbitrary λ. In the formula (6) we ignore the second term on the right hand side for the reason that it is significantly smaller then the first term. Indeed, it appeared that the unbiased estimate of the second term in (6) is more then 40 times smaller then that of the first term. Taking the second term into account will significantly complicate calculations but will not change our results noticeably. The complications result from the fact that it is not straightforward to construct an unbiased estimate of this term. Presence of several sums in the expression of the estimate of δ will create unnecessary difficulties in computing its variance.

We propose the following unbiased estimate for δ:

$$\tilde{\delta} = \frac{1-\lambda}{\lambda}\frac{1}{T}\sum_t \tilde{\pi}_{t-1}(\tilde{p}_t^{s|s} - \tilde{\pi}_t) = \frac{1-\lambda}{\lambda}\tilde{a} \tag{8}$$

The absence of a bias stems from the fact that estimates are independent in time. Variance of a can be easily estimated the following way:

$$V(\tilde{a}) = \frac{1}{T^2}\sum_t V(\tilde{\pi}_{t-1}(\tilde{p}_t^{s|s} - \tilde{\pi}_t)) - \frac{1}{T^2}\sum_t \pi_{t-2}(p_t^{s|s} - \pi_t)(Cov(\tilde{\pi}_{t-1}, \tilde{p}_{t-1}^{s|s}) - V(\tilde{\pi}_{t-1})) \tag{12}$$

where

$$V(\tilde{\pi}_{t-1}(\tilde{p}_t^{s|s} - \tilde{\pi}_t)) = E[\tilde{\pi}_{t-1}^2]E[\tilde{p}_t^{s|s} - \tilde{\pi}_t]^2 - \pi_{t-1}(p_t^{s\}s} - \pi_t) \tag{13}$$

Here we gain use independence of estimates in time. Expressions (12) and (13) include terms $V(\tilde{\pi}_t)$, $Cov(\tilde{\pi}_t, \tilde{p}_t^{s|s})$, $E[\tilde{p}_t^{s|s} - \tilde{\pi}_t]^2$ and $E[\tilde{\pi}_t]^2$. Let us find explicit formula for them.

Let N_t^{ss} be the number of investors that succeeded yesterday and today, N_t^{fs} – number of investors that failed yesterday but succeeded today, N_t^{sf} – number of investors that succeeded today but failed yesterday and finally N_t^{ff} – number of investor that failed both today and yesterday. Define $N_t^s = N_t^{ss} + N_t^{fs}$ – number of investors that succeeded yesterday and $N_t^f = N_t^{sf} + N_t^{ff}$ – number of investors that failed yesterday. By definition $\tilde{p}_t^{s|s} = \frac{N_t^{ss}}{N_t^s}$ and $\tilde{p}_t^{f|s} = \frac{N_t^{sf}}{N_t^f}$ are two independent random variables that have means $p_t^{s|s}$ and $p_t^{s|f}$ correspondingly. The frequency of success is a linear combination of them:

$$\tilde{\pi}_t = \frac{\tilde{p}_t^{s|s}N_t^s + \tilde{p}_t^{s|f}N_t^f}{N_t^s + N_t^f}$$

Hence we have:

$$Cov(\tilde{\pi}_t, \tilde{p}_t^{s|s}) = \frac{N_t^s}{N_t^s + N_t^f}V(\tilde{p}_t^{s|s})$$

Assuming normality of estimates, we can easily find the following expressions for variances:

$$V(\tilde{p}_t^{s|s}) = \frac{p_t^{s|s}(1 - p_t^{s|s})}{N_t^s}$$

$$V(\tilde{\pi}_t) = \frac{\pi_t(1-\pi_t)}{N_t^s + N_t^f}$$

From the definition of the variance:

$$E[\tilde{\pi}_t^2] = V(\tilde{\pi}_t) + \pi_t^2$$

$$E[\tilde{p}_t^{s|s} - \tilde{\pi}_t]^2 = V(\tilde{p}_t^{s|s}) + V(\tilde{\pi}_t) - 2Cov(\tilde{p}_t^{s|s}, \tilde{\pi}_t) - (p_t^{s|s} - \pi_t)^2$$

Variance $V(\tilde{a})$ can be evaluated by substituting true probabilities with their unbiased estimates. In order to find confidence intervals we assume that $\tilde{a}$ is normally distributed. Indeed, as it follows from (8) it is an average of random variables with T being equal to 240. The normality assumption is justified by the Central limit theorem.

Let us assume that computed estimate of π is exact. Indeed, the standard error of the estimate is very small (0.0035) so this assumption will not affect our results in a significant way. Taking into account (7) we conclude that given λ and estimate $\tilde{a}$, the true value p^h belongs to the following interval with 95% probability:

$$\pi + \sqrt{\frac{1-\lambda}{\lambda} + t\frac{1-\lambda}{\lambda}\sqrt{\tilde{V}(\tilde{a})}} \quad \text{where } t \in [-2,2] \tag{14}$$

Now using (4) and (14) we find that the pair of true parameters (p^h, p^l) with al least 95% probability belongs to a subset $\{(x,y)\}$ of R^2 that can be parameterized in the following way:

$$x = \pi + \sqrt{\frac{1-\lambda}{\lambda}\tilde{a} + t\frac{1-\lambda}{\lambda}\sqrt{\tilde{V}(\tilde{a})}} \tag{15}$$

$$y = \frac{\pi - \lambda x}{1-\lambda} \tag{16}$$

$$t \in [-2,2],\ \lambda \in (0,1), \text{ such that } (x,y) \in [0,1]^2 \tag{17}$$

This set represents two-dimensional 95% confidence area, which is consistent with our observation. The dimension two results from the fact that we do not observe λ, and it is difficult to make any prior judgment about it.